Stochastic Processes and their Applications 16 (1983) 71–84
North-Holland

INVARIANCE PRINCIPLES FOR STOCHASTIC AREA AND RELATED STOCHASTIC INTEGRALS

Svante JANSON* and Michael J. WICHURA**
Departments of Mathematics and Statistics, University of Chicago, Chicago, IL 60637, USA

Received 6 April 1981

Given an antisymmetric kernel K ($K(z,z')=-K(z',z)$) and i.i.d. random variates Z_n, $n\geq 1$, such that $\mathrm{E}\,K^2(Z_1,Z_2)<\infty$, set $A_n=\sum_{1\leq i<j\leq n}K(Z_i,Z_j)$, $n\geq 1$. If the Z_n's are two-dimensional and K is the determinant function, A_n is a discrete analogue of Paul Lévy's so-called stochastic area. Using a general functional central limit theorem for stochastic integrals, we obtain limit theorems for the A_n's which mirror the corresponding results for the symmetric kernels that figure in theory of U-statistics.

AMS (MOS) Subj. Class.: Primary 60F05; Secondary 60H65, 60J65
stochastic integral — invariance principle
stochastic area — antisymmetric kernel

1. A functional central limit theorem for stochastic integrals

For each n, let $(S_{nk})_{1\leq k\leq k_n}$ be a mean 0 square integrable martingale adapted to increasing σ-fields $(\mathscr{F}_{nk})_{0\leq k\leq k_n}$, and let $(\sigma_{nk})_{1\leq k\leq k_n}$ be a system of (perhaps degenerate) random variables such that

$$0=\sigma_{n0}<\sigma_{n1}<\cdots<\sigma_{nk_n}=1$$

and such that σ_{nk} is $\mathscr{F}_{n,k-1}$ measurable for each k. Let $W_n=(W_n(t))_{0\leq t\leq 1}$ be the random step function with $W_n(0)=0$ and with jumps $X_{nk}=S_{nk}-S_{n,k-1}$ at the points σ_{nk}:

$$W_n(t)=\sum_{\sigma_{nk}\leq t}X_{nk}=S_{n,\max\{k:\sigma_{nk}\leq t\}}.$$

Suppose that W_n converges weakly to a standard Wiener process W with respect to the Skorohod topology on the space $D=D([0,1],\mathbb{R})$; write this as

$$W_n\to_{\mathrm{D}} W. \tag{1.1}$$

* Present address: Department of Mathematics, Uppsala University, Thunbersvägen 3, S-75238 Uppsala, Sweden.

** Support for this research was provided in part by National Science Foundation Grant No. MCS 80-02217.

Under a mild supplementary condition, we establish the weak convergence

$$J_n = \int_0^{\cdot} F(W_n)\,\mathrm{d}W_n \to_{\mathrm{D}} J = \int_0^{\cdot} F(W)\,\mathrm{d}W \tag{1.2}$$

of the corresponding stochastic integrals of a W-continuous nonanticipative map $F: D \to D$.

To be more specific, the hypotheses on F are these:

F is measurable between $\mathscr{C}$ and $\mathscr{C}$,

where $\mathscr{C}$ is simultaneously the Borel σ-field of the Skorohod topology and the coordinate σ-field $\sigma\langle \pi_t : 0 \leq t \leq 1\rangle$ of D,

$\{x \in D : F$ is continuous at $x\}$ has probability one under W,

and

$$(F(x))(t) = (F(y))(t)$$

for any $x, y \in D$ and $t \in [0, 1]$ with

$$x(s) = y(s) \quad \text{for all } s \leq t.$$

We take

$$J_n(t) = \int_0^t FW_n\,\mathrm{d}W_n = \lim_{\{t_i\}} \sum_i (FW_n)(t_i)(W_n(t_{i+1}) - W_n(t_i))$$

$$= \sum_{\sigma_{nk} \leq t} (FW_n)(\sigma_{nk}-)X_{nk}$$

where the (pointwise) limit is taken over finite systems

$$0 = t_0 < t_1 < \cdots < t_m = t$$

as $\max_{l \leq m}(t_l - t_{l-1}) \to 0$. Evidently J_n is a random element of D. In addition we take

$$J(t) = \int_0^t FW\,\mathrm{d}W = \lim_{\{t_i\}} \sum_i (FW)(t_i)(W(t_{i+1}) - W(t_i))$$

to be the corresponding Ito integral, which is defined since FW has bounded sample paths and since for each t

$$FW(t) = F((W_{\min(s,t)})_{0 \leq s \leq 1})(t)$$

is measurable with respect to the pre-t σ-field $\sigma\langle W_s : s \leq t\rangle$ of W; the limit defining $J(t)$ exists in probability uniformly in t, and J (by convention) has continuous sample paths. Although (1.2) looks like an application of the mapping principle to (1.1), it is not, since the function

$$x \to \int_0^{\cdot} F(x)\,\mathrm{d}x$$

is not even defined pointwise for relevant $x \in D$.

For the condition to supplement (1.1) we assume either

Condition A. FW has continuous sample paths, and

$$\lim_{c\to\infty} \limsup_n \mathrm{P}\Big\{\sum_k E_{n,k-1}X_{nk}^2 \geq c\Big\} = 0,$$

with E_{nj} denoting conditional expectation given $\mathscr{F}_{nj}$, or

Condition B. For each t in [0, 1],

$$\sum_{\sigma_{nk}\leq t} E_{n,k-1}X_{nk}^2 \to_{\mathrm{P}} t.$$

For circumstances under which (1.1) holds along with one of these conditions, consult e.g. [1, 3, 4, 5, 6, 7, 11, 12, 16, 17, 18, 20, 22, 23].

Theorem 1.1. *Under the assumptions made above,* (1.2) *holds.*

Proof. Write V_n for FW_n, $n \geq 1$, and V for FW. There are at most countably many points t of fixed discontinuity ($\mathrm{P}\{V(t) \neq V(t-)\} > 0$) of V. Since the values our processes take at $t=1$ do not figure into any of the stochastic integrals, we may and do assume that 1 is not a fixed discontinuity of V. For convenience we further assume that in fact no rationals fall among these points; if this is not the case, the grids $(p/q)_{0\leq p\leq q}$ utilized below will have to be jiggled a tiny bit to avoid the fixed dicontinuity points.

For each integer $q \geq 1$, let $A_q\colon D \to D$ be defined by

$$x_q(t) \equiv (A_q(x))(t) = x\left(\frac{[tq]}{q}\right)$$

($x \in D$, $0 \leq t \leq 1$). Set

$$F_q = A_q F,$$

$$V_{nq} = A_q V_n = A_q FW_n = F_q W_n, \qquad V_q = A_q V = A_q FW = F_q W,$$

$$J_{nq}^* = \int_0^{\cdot} V_{nq}\,\mathrm{d}W_n, \qquad J_q^* = \int_0^{\cdot} V_q\,\mathrm{d}W.$$

Notice that F_q is W-continuous and nonanticipative. The plan is to show

$$J_{nq}^* \to_{\mathrm{D}} J_q^* \quad \text{as } n\to\infty, \tag{1.3}$$

$$J_q^* \to_{\mathrm{D}} J \quad \text{as } q\to\infty \tag{1.4}$$

and

$$\operatorname*{plim}_{q\to\infty}\lim_{n\to\infty} \|J_{nq}^* - J_n\| = 0 \tag{1.5}$$

in the sense that

$$\lim_q \limsup_n \mathrm{P}\{\sup_t |J_{nq}^*(t) - J_n(t)| \geq \varepsilon\} = 0$$

for each $\varepsilon > 0$, and to deduce (1.2) from Billingsley's triangle theorem (Theorem 4.2, Billingsley [2]).

In regard to (1.3), note

$$J_{nq}^*(t) = \int_0^t V_{nq}\,\mathrm{d}W_n = \sum_{\sigma_{nk} \le t} V_{nq}(\sigma_{nk}-)X_{nk}$$
$$= V_n(0)W_n\left(\frac{1}{q}\right) + V_n\left(\frac{1}{q}\right)\left(W_n\left(\frac{2}{q}\right) - W_n\left(\frac{1}{q}\right)\right)$$
$$+ \cdots + V_n\left(\frac{p}{q}\right)\left(W_n(t) - W_n\left(\frac{p}{q}\right)\right)$$

for $p/q \le t \le (p+1)/q$. V and W being continuous at multiples of $1/q$, (1.3) follows from (1.1) and the mapping theorem for weak convergence. As for (1.4), since each $x \in D$ is bounded and continuous at all but at most countably many t,

$$\int (V_q(t) - V(t))^2\,\mathrm{d}t = \int_0^1 \left(V\left(\frac{[tq]}{q}\right) - V(t)\right)^2 \mathrm{d}t \to 0$$

pointwise, so

$$\left\|\int_0^{\cdot} V_q\,\mathrm{d}W - \int_0^{\cdot} V\,\mathrm{d}W\right\| \to_{\mathrm{P}} 0$$

by the mapping theorem for Ito integrals (Theorem 3.4, Friedman [8]).

Consider now (1.5). One has

$$\|J_{nq}^* - J_n\| = \max_l |Y_{nl}|$$

where

$$Y_{nl} = \sum_{k \le l} C_{nkq} X_{nk}$$

with

$$C_{nkq} = V_n(\sigma_{nk}-) - V_{nq}(\sigma_{nk}-).$$

Because F is nonanticipative, $V_n(\sigma_{nk}-) = (FW_n)(\sigma_{nk}-)$ depends in a measurable way on X_{nj} and σ_{nj}, for $j < k$, and on σ_{nk}; the same considerations apply to $V_{nq}(\sigma_{nk}-)$, so C_{nkq} is $\mathscr{F}_{n,k-1}$ measurable for each k. Apply Kolmogorov's inequality to the submartingale $(\sum_{k \le l} C_{nkq}^* X_{nk})_{l \ge 1}^2$, with

$$C_{nkq}^* = \begin{cases} C_{nkq} & \text{if } \sum_{j \le k} C_{njq}^2 E_{n,j-1} X_{nj}^2 \le \rho\varepsilon^2, \\ 0 & \text{otherwise,} \end{cases}$$

to get

$$\mathrm{P}\{\max_l |Y_{nl}| \ge \varepsilon\} \le \mathrm{P}\left\{\sum_k C_{nkq}^2 E_{n,k-1} X_{nk}^2 > \rho\varepsilon^2\right\} + \rho$$

for each $\varepsilon > 0$ and $\rho > 0$. It suffices then to show

$$\operatorname*{plim}_q \lim_n \sum_k C_{nkq}^2 E_{n,k-1} X_{nk}^2 = 0. \tag{1.6}$$

To argue (1.6) under Condition A, note that

$$\max_k |C_{nkq}| \leqslant \omega_{1/q}(V_n),$$

where

$$\omega_\delta(x) = \sup\{|x(t) - x(s)| : 0 \leqslant s < t \leqslant 1, t - s \leqslant \delta\}$$

is the usual modulus of uniform continuity. Given that V has continuous sample paths,

$$\omega_\delta(V_n) \to_{\mathrm{D}} \omega_\delta(V)$$

as $n \to \infty$, and

$$\omega_\delta(V) \to 0$$

as $\delta \to 0$, whence the stochastic boundedness of $\sum_k E_{n,k-1}X_{nk}^2$ implies (1.6).

To argue (1.6) under Condition B, let λ_n be the (random) measure on $[0, 1]$ placing masses $E_{n,k-1}X_{nk}^2$ at the points σ_{nk}, and observe that

$$\sum_k C_{nkq}^2 E_{n,k-1} X_{nk}^2 = \int_0^1 (V_n(t-) - V_{nq}(t-))^2 \lambda_n(\mathrm{d}t).$$

It will be enough to show

$$\int_0^1 (V_n(t-) - V_{nq}(t-))^2 \lambda_n(\mathrm{d}t) \to_{\mathrm{D}} \int_0^1 (V(t) - V_q(t))^2 \,\mathrm{d}t$$

as $n \to \infty$, for, as already noted, the integral on the right-hand side above tends to 0 as $q \to \infty$. Now by hypothesis,

$$\lambda_n \to_{\mathrm{P}} \lambda$$

where λ (nonrandom) is Lebesgue measure on $[0, 1]$, so

$$(V_n, \lambda_n) \to_{\mathrm{D}} (V, \lambda)$$

(see Theorem 4.4, Billingsley [2]). By the extended mapping theorem (Theorem 5.5, Billingsley [2]), it is enough to check that

$$x_n \to x \text{ in } D \quad \text{and} \quad \mu_n \to \lambda,$$

x_n, μ_n $(n \geqslant 1)$, and x being nonrandom, and x being continuous at multiples of $1/q$, entails

$$\int_0^1 (x_n(t-) - x_{nq}(t-))^2 \mu_n(\mathrm{d}t) \to \int_0^1 (x(t) - x_q(t))^2 \,\mathrm{d}t.$$

But this itself results from the extended mapping theorem after assuming, as one may, that μ_n is a probability measure, and noting that

$$x_n(t_n-) - x_{nq}(t_n-) \to x(t) - x_q(t)$$

whenever $t_n \to t$ and x is continuous at $t \not\equiv 0 \bmod(1/q)$, the convergence being in any case uniformly bounded. □

As is clear from the proof, various extensions of Theorem 1.1 may be made (we have put off mentioning them until now in order not to obscure the argument), including the following:

(i) the S_{nk}'s and $W(t)$'s may take values in $\mathbb{R}^r$ for some r, in which case the stochastic integrals are defined as usual using the dot product of the integrand and the integrator, and Conditions A and B are applied to each of the r marginal processes,

(ii) $W(1)$ may have an arbitrary covariance structure, in which case the marginal variances will have to be figured into Condition B,

(iii) given several F's, (1.2) holds for them jointly,

(iv) F may depend on n, provided

$$\{x : x_n \to x \text{ implies } F_n(x_n) \to F(x)\}$$

has W-probability 1,

(v) the k_n's may be random, and

(vi) if there are infinitely many variates S_{nk} in each row, to be plotted at arbitrarily large times, the whole affair may be recast in $D([0, \infty))$.

We will use (i) in the next section.

2. Stochastic area

Let $W = (W(s))_{s \geq 0}$, with $W(s) = \binom{U(s)}{V(s)}$, be a standard two-dimensional Brownian motion emanating from the origin. During the instant s to $s + ds$, the chord from the origin to W sweeps out a triangular region of area

$$\tfrac{1}{2}R(s)\,dN(s),$$

where

$$R(s) = \sqrt{(U^2(s) + V^2(s))}$$

and

$$dN(s) = -\frac{V(s)}{R(s)}\,dU(s) + \frac{U(s)}{R(s)}\,dV(s),$$

so the Ito integral

$$A(t) = \int_0^t R(s)\,dN(s) = \int_0^t -V(s)\,dU(s) + U(s)\,dV(s)$$

records twice the area swept out in the first t units of time. Paul Lévy ([14, 15]) studied the so-called *stochastic area* process $A = (A(t))_{t \geq 0}$ and found that $A(t)$ has characteristic function

$$E\,e^{i\theta A(t)} = E\,e^{i\theta t A(1)} = \frac{1}{\cosh(\theta t)},$$

$-\infty < \theta < \infty$, and density

$$f_{A(t)}(a) = \frac{1}{2t\cosh(\pi a/2t)},$$

$-\infty < a < \infty$. $\boldsymbol{A}$ obeys the generalized reflection principle, in that the process obtained from $\boldsymbol{A}$ by changing the sign of the increments of $\boldsymbol{A}$ after a stopping time of $\boldsymbol{W}$ is probabilistically indistinguishable from $\boldsymbol{A}$ (indeed, the Brownian motion $\boldsymbol{N}$ has this property, and $\boldsymbol{N}$ is independent of $\boldsymbol{R}$). Accordingly the distribution of

$$m(t) = \inf_{0 \le s \le t} A(s), \quad M(t) = \sup_{0 \le s \le t} A(s) \quad \text{and} \quad A(t)$$

can be written down simply in terms of the distribution of $\boldsymbol{A}(t)$; for example

$$M(t) =_{\mathrm{D}} |A(t)|.$$

Now let $Z_n = \binom{X_n}{Y_n}$, $n \ge 1$, be i.i.d. random vectors, with $E(Z_n) = 0$ and $\mathrm{Cov}(Z_n) = I^{2\times 2}$. As an analogue to $\boldsymbol{A}$ consider the *discrete stochastic area* process $A = (A_n)_{n \ge 1}$ with

$$A_n = \sum_{1 \le i < j \le n} (-Y_i X_j + X_i Y_j).$$

Sums of this type occur in physics in connection with the Heisenberg group $\{[x, y, z] \in \mathbb{R}^3\}$ with multiplication

$$[x, y, z] * [x', y', z'] = [x + x', y + y', z + z' - yx' + xy'];$$

indeed

$$\prod_{i=1}^{n} [X_i, Y_i, 0] = \left[\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i, A_n\right].$$

Since

$$\frac{A_m}{n} = \sum_{1 \le j \le m} \left[-\frac{T_{j-1}}{\sqrt{n}}\frac{X_j}{\sqrt{n}} + \frac{S_{j-1}}{\sqrt{n}}\frac{Y_j}{\sqrt{n}}\right],$$

where

$$S_k = \sum_{i \le k} X_i \quad \text{and} \quad T_k = \sum_{i \le k} Y_i,$$

and since

$$\frac{1}{\sqrt{n}} Z_{[n\cdot]} = \frac{1}{\sqrt{n}} \binom{S_{[n\cdot]}}{T_{[n\cdot]}} \to_{\mathrm{D}} \binom{U}{V} = W$$

by the two-dimensional Donsker theorem, we obtain from Theorem 1.1 (with $F: D([0, 1], \mathbb{R}^2) \to D([0, 1], \mathbb{R}^2)$ defined by $F\binom{x}{y} = \binom{-y}{x}$) a functional central limit theorem for stochastic area:

$$(A_m(t))_{0\le t\le 1} \equiv \left(\frac{A([nt])}{n}\right)_{0\le t\le 1} \to_{\mathrm{D}} (\boldsymbol{A}(t))_{0\le t\le 1}. \tag{2.1}$$

In particular

$$\frac{A_n}{n} \to_{\mathrm{D}} \boldsymbol{A}(1),$$

as was obtained by Guivarc'h, Keane, and Roynette [10, p. 151].

The validity of (2.1) would follow from Theorem 1.1 even for certain martingales. We consider here a different sort of generalization of (2.1) based on the observation that

$$A_n = \sum_{1\le i<j\le n} K(Z_i, Z_j),$$

where

$$K\left(\binom{x}{y}, \binom{x'}{y'}\right) = \det\begin{pmatrix} x & x' \\ y & y' \end{pmatrix} = xy' - yx' = -K\left(\binom{x'}{y'}, \binom{x}{y}\right)$$

is an antisymmetric function of its arguments and is square integrable with respect to the joint distribution of Z_1 and Z_2. In what follows we present an extension of (2.1) to the case of a general antisymmetric square integrable kernel K. The results parallel those for the symmetric square integrable kernels that figure in the theory of U-statistics.

Throughout the remaining discussion, suppose that $(\boldsymbol{Z}, \mathscr{B}, \mu)$ is an arbitrary probability space, and let $K: \boldsymbol{Z} \times \boldsymbol{Z} \to \mathbb{R}$ be antisymmetric:

$$K(z, z') = -K(z', z),$$

and square integrable:

$$\int K^2(z, z')\, \mu(\mathrm{d}z)\, \mu(\mathrm{d}z') < \infty.$$

Antisymmetry forces

$$\int K(z, z')\, \mu(\mathrm{d}z)\, \mu(\mathrm{d}z') = 0.$$

Lemma 2.1. *K admits an expansion as a finite or infinite series*

$$K(z, \cdot\,) = \sum_r \lambda_r(g_r(z)h_r(z') - h_r(z)g_r(z')), \tag{2.2}$$

convergent in $L_2(\mu \times \mu)$, where the g_r's and h_r's are orthonormal in $L_2(\mu)$, and where the λ_r's are positive numbers with

$$2\sum_r \lambda_r^2 = \int K^2(z, z')\, \mu(\mathrm{d}z)\, \mu(\mathrm{d}z') < \infty.$$

Proof. For complex f in $L_2(\mu)$ define

$$Tf(z) = \int K(z, z') f(z')\, \mu(\mathrm{d}z').$$

T is a Hilbert–Schmidt operator on $L_2(\mu)$, and $\mathrm{i}T$ is selfadjoint, so in $L_2(\mu \times \mu)$ we may write

$$K(z, z') = \sum_r \mathrm{i}\lambda_r f_r(z) \bar{f}_r(z')$$

where the f_r's are orthonormal eigenvectors of T with associated nonzero eigenvalues $\mathrm{i}\lambda_r$, λ_r real. If f is an eigenvector, then so is $\bar{f}$, and the corresponding eigenvalues are conjugate. Hence the expansion becomes

$$\begin{aligned} K(z, z') &= \sum_{\lambda_r > 0} \mathrm{i}\lambda_r (f_r(z)\bar{f}_r(z') - \bar{f}_r(z) f_r(z')) \\ &= \sum_{\lambda_r > 0} 2\lambda_r (g_r(z) h_r(z') - h_r(z) g_r(z')) \end{aligned}$$

where

$$f_r = g_r + \mathrm{i}h_r \quad \text{and} \quad \bar{f}_r = g_r - \mathrm{i}h_r$$

with f_r and g_r real. From the orthonormality of the various f_r and $\bar{f}_r$ follows the orthogonality of the various g_r and h_r as well as

$$\int g_r^2\, \mathrm{d}\mu = \tfrac{1}{2} = \int h_r^2\, \mathrm{d}\mu.$$

Renormalization gives (2.2). □

Set now

$$\kappa(z) = \int K(z, z')\, \mu(\mathrm{d}z') = \sum_r \lambda_r (\eta_r g_r(z) - \gamma_r h_r(z))$$

(holding in $L_2(\mu)$), where

$$\gamma_r = \int g_r(z)\, \mu(\mathrm{d}z) \quad \text{and} \quad \eta_r = \int h_r(z)\, \mu(\mathrm{d}z),$$

and put

$$\sigma^2 = \int \kappa^2(z)\, \mu(\mathrm{d}z) = \sum_r \lambda_r^2 (\gamma_r^2 + \eta_r^2).$$

Theorem 2.1. *Let $Z, Z_1, Z_2, \ldots$ be i.i.d. Z-random variables with distribution μ, and set*

$$A_n = \sum_{1 \le i < j \le n} K(Z_i, Z_j).$$

Degenerate case. *If*

$$\sigma^2 \equiv E\kappa^2(Z) = 0$$

then

$$\left(\frac{A_{[nt]}}{n}\right)_{0 \le t \le 1} \to_{\mathrm{D}} \left(\sum_r \lambda_r \mathbf{A}_r(t)\right)_{0 \le t \le 1} \tag{2.3}$$

where the $\mathbf{A}_r$'s are independent stochastic area processes, and the sum on the right-hand side of (2.3) converges uniformly in t, with probability one.

Nondegenerate case. *If on the other hand*

$$\sigma^2 > 0,$$

then

$$\left(\frac{1}{\sqrt{n}} \frac{A_{[nt]}}{[nt]}\right)_{0 \le t \le 1} \to_{\mathrm{D}} \sigma\left(W(t) - \frac{2}{t}\int_0^t W(s)\,\mathrm{d}s\right)_{0 \le t \le 1} \tag{2.4}$$

where W is a standard Brownian motion process.

Comparable results for symmetric square integrable K (with $\mathrm{E}\,K(Z_1, Z_2) = 0$) are well known in the theory of U-statistics; in the degenerate case, Neuhaus [21] has (2.3) with the right-hand side replaced by

$$\left(\sum_r \theta_r \int_0^t W_r(s)\,\mathrm{d}W_r(s)\right)_{0 \le t \le 1} = \left(\tfrac{1}{2}\sum_r \theta_r (W_r^2(t) - t)\right)_{0 \le t \le 1}, \tag{2.5}$$

where the θ_r's are the eigenvalues of K (symmetric) and the W_r's are independent standard Brownian motions; while in the nondegenerate case, Miller and Sen [19] have (2.4) with the right hand side replaced simply by σW. As in Hall's [13] treatment of the symmetric case, it is not hard to present a single result that contains both (2.3) and (2.4), but we have chosen to forgo this for the sake of simplicity. For ease of reference, call $(A_n)_{n \ge 1}$ the *discrete stochastic K-area process.*

Proof. *The degenerate case.* Here the variates $g_r(Z)$ and $h_r(Z)$, $r \ge 1$, have zero means in addition to being orthogonal. For integers $1 \le q < s \le \infty$, write

$$K_{q,s}(z, z') = \sum_{q < r \le s} \lambda_r (g_r(z) h_r(z') - h_r(z) g_r(z')),$$

$$A_{n;q,s} = \sum_{1 \le i < j \le n} K_{q,s}(Z_i, Z_j),$$

$$Y_{n;q,s} = \frac{1}{n}(A_{[nt];q,s})_{0 \le t \le 1}.$$

For $s<\infty$ it follows from Donsker's theorem in $\prod_{r=q}^{s}\mathbb{R}^2$ and the considerations used for (2.1) that

$$Y_{n;q,s} \to_{\mathrm{D}} \sum_{r=q}^{s} \lambda_r \boldsymbol{A}_r, \tag{2.6}$$

where $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$ are independent stochastic area processes on $[0, 1]$.

Now for any $q<s$,

$$\Delta_n \equiv A_{n;q,s} - A_{n-1;q,s} = \sum_{1\le i<n} K_{q,s}(Z_i, Z_n)$$

is a martingale difference sequence, so by Kolmogorov's inequality

$$\begin{aligned}
\mathrm{P}\{\|Y_{n;q,s}\| \ge c\} &= \mathrm{P}\{\max_{m\le n} |A_{m;q,s}| \ge cn\} \\
&\le \frac{1}{c^2n^2}\mathrm{E}(A^2_{n;q,s}) \\
&= \frac{1}{c^2n^2} \sum_{1\le i<j\le n}\ \sum_{1\le k<l\le n} \mathrm{E}(K_{q,s}(Z_i, Z_j)K_{q,s}(Z_k, Z_l)) \\
&= \frac{1}{c^2n^2} \sum_{1\le i<j\le n} \mathrm{E}\, K^2_{q,s}(Z_i, Z_j) \\
&\le \frac{1}{2c^2}\mathrm{E}\, K^2_{q,s} = \frac{1}{c^2}\Big(\sum_{r=q}^{s} \lambda_r^2\Big),
\end{aligned}$$

for $c>0$. Hence

$$\operatorname*{plim}_{s\to\infty}\lim_{n\to\infty} \|Y_{n;1,\infty} - Y_{n;1,s}\| = \operatorname*{plim}_{s}\lim_{n} \|Y_{n;s+1,\infty}\| = 0. \tag{2.7}$$

Moreover, because of (2.6),

$$\operatorname*{plim}_{q,s\to\infty} \Big\| \sum_{r=q}^{s} \lambda_r \boldsymbol{A}_r \Big\| = 0, \tag{2.8}$$

so $\sum_r \lambda_r \boldsymbol{A}_r$ converges uniformly in probability, and even with probability one since the summands are independent. Eq. (2.3) now follows from (2.6), (2.7) and (2.8) via Billingsley's triangle theorem.

The nondegenerate case. Here $\kappa(Z)$ has mean 0 and, crucially, variance $\sigma^2 = \sum_r \lambda_r(\gamma_r^2+\eta_r^2)>0$. Write

$$K(z, z') = K^*(z, z') + \kappa(z) - \kappa(z')$$

and observe that

$$\kappa^*(z) \equiv \mathrm{E}(K^*(z, Z)) = 0$$

for each z, so K^* falls into the degenerate case. Writing X_i for $\kappa(Z_i)$ we have

$$\begin{aligned} A_n &= \sum_{1\le i<j\le n} K(Z_i, Z_j) \\ &= \sum_{1\le i<j\le n} K^*(Z_i, Z_j) + \sum_{1\le i<j\le n} X_i - \sum_{1\le i<j\le n} X_j \\ &\equiv A_n^* + \sum_{i=1}^{n} (n-2i+1)X_i. \end{aligned}$$

To obtain (2.4) it suffices to show that

$$\frac{1}{\sqrt{n}} \max_{m\le n} \frac{|A_m^*|}{m} \to_{\mathrm{P}} 0 \tag{2.9}$$

and that

$$V_n \to_{\mathrm{D}} V, \tag{2.10}$$

where

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \left(1 - \frac{2i}{[nt]} + \frac{1}{[nt]}\right) X_i$$

and

$$V(t) = \sigma \int_0^t \left(1 - 2\frac{s}{t}\right) \mathrm{d}W(s) = \sigma\left(-W(t) + \frac{2}{t}\int_0^t W(s)\,\mathrm{d}s\right),$$

$0 \le t \le 1$.

Now (2.9) is immediate from the martingale properties of A_n^* and the Hájek–Renyi inequality:

$$\begin{aligned} \mathrm{P}\{|A_m^*| \ge cm\sqrt{n} \text{ for some } m\} &\le \frac{1}{c^2 n} \sum_{m=1}^{n} \frac{1}{m^2} \mathrm{E}\Big(\sum_{1\le i<m} K^*(Z_i, Z_m)\Big)^2 \\ &= \frac{1}{c^2 n} \sum_{m=1}^{n} \frac{m-1}{m^2} \mathrm{E}(K^*)^2 \\ &= \mathrm{O}\left(\frac{\log n}{n}\right). \end{aligned}$$

Moreover, because

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{1\le i\le m} \left(1 - \frac{2i}{m} + \frac{1}{m}\right) X_i &= \frac{1}{\sqrt{n}}\left[-\left(1-\frac{1}{m}\right) S_m + \frac{2}{m}\sum_{1\le i<m} S_i\right], \\ &= -\left(1-\frac{1}{m}\right) W_n\left(\frac{m}{n}\right) + 2\frac{n}{m}\frac{1}{n}\sum_{1\le i<m} W_n\left(\frac{i}{n}\right), \end{aligned}$$

where

$$S_i = X_1 + \cdots + X_i \quad \text{and} \quad W_n = \frac{S_{[n\cdot]}}{\sqrt{n}},$$

(2.10) follows from Donsker's theorem and the extended mapping theorem. □

Our methods apply equally well to the general kernel K satisfying $\mathrm{E}\, K^2(Z_1, Z_2) < \infty$ and $\mathrm{E}\, K(Z_1, Z_2) = 0$, and we close with a few remarks concerning the extension of Theorem 2.1 to this more general situation. In the nondegenerate case

$$\max_{i=1,2} \operatorname{Var} \mathrm{E}(K(Z_1, Z_2)|Z_i) > 0,$$

one gets in place of (2.4) that

$$\left(\frac{1}{\sqrt{n}} \frac{A_{[nt]}}{[nt]}\right)_{0 \le t \le 1} \to_{\mathrm{D}} \left(W'(t) + \frac{1}{t} \int_0^t (W(s) - W'(s))\, \mathrm{d}s\right)_{0 \le t \le 1}, \tag{2.11}$$

where $\binom{W}{W'}$ is a two-dimensional Brownian motion with the same covariance matrix as $\binom{\kappa(Z)}{\kappa'(Z)}$, with

$$\kappa(z) = \mathrm{E}\, K(z, Z) \quad \text{and} \quad \kappa'(z') = \mathrm{E}\, K(Z, z').$$

In the degenerate case,

$$\begin{aligned} K(z, z') &= \tfrac{1}{2} K_s(z, z') + \tfrac{1}{2} K_a(z, z') \\ &\equiv \tfrac{1}{2}(K(z, z') + K(z', z)) + \tfrac{1}{2}(K(z, z') - K(z', z)) \end{aligned}$$

decomposes as the sum of a symmetric and an antisymmetric kernel, each of which is degenerate, and one gets in place of (2.3) that

$$2\left(\frac{A_{[nt]}}{n}\right)_{0 \le t \le 1} \to_{\mathrm{D}} S + D, \tag{2.12}$$

where S is of the form (2.5), with the θ_r's being the eigenvalues of K_s, and where D has the same structure as the process on the right-hand side of (2.3), with the λ_r's being the eigenvalues of K_a. S and D are not in general independent, due to the possible nonorthogonality of the eigenvectors of K_s from those of K_a. For example, in the simple case $Z = \binom{X}{Y}$ with $E(Z) = 0$, $\operatorname{Cov}(Z) = I^{2\times 2}$, and $K(\binom{x}{y}, \binom{x'}{y'}) = xy'$, (2.12) holds with

$$S(t) = \int_0^t U\, \mathrm{d}V + V\, \mathrm{d}U = U(t)V(t)$$

and

$$D(t) = \int_0^t U\, \mathrm{d}V - V\, \mathrm{d}U = A(t),$$

U and V being independent standard Brownian motions, and by calculation one obtains

$$\mathrm{E}\, \mathrm{e}^{\mathrm{i}(\rho S(1) + \theta D(1))} = \left(\cosh^2(\theta) + \left(\frac{\rho}{\theta}\right)^2 \sinh^2(\theta)\right)^{-1/2}.$$

Note the curious fact that the characteristic function of $D(1)$ is the product of those of $\int_0^1 U\, \mathrm{d}V$ and $-\int_0^1 V\, \mathrm{d}U$, even though these two variates are not independent.

References

[1] D. Aldous, Stopping times and tightness, Ann. Probab. 6 (1978) 335–340.
[2] P. Billingsley, Convergence of Probability Measures (Wiley, New York, 1968).
[3] D. Blackwell and D.A. Freedman, On the amount of variance needed to escape from a strip, Ann. Probab. 1 (1973) 772–787.
[4] B.M. Brown, Martingale central limit theorems, Ann. Math. Statist. 42 (1971) 59–66.
[5] R. Drogin, An invariance principle for martingales, Ann. Math. Statist. 43 (1972) 602–620.
[6] D.A. Freedman, Brownian Motion and Diffusion (Holden-Day, San Francisco, 1971).
[7] D.A. Freedman, On tail probabilities for martingales, Ann. Probab. 3 (1975) 100–118.
[8] A. Friedman, Stochastic Differential Equations, Vol. I (Academic Press, New York, 1975).
[9] G.C. Gregory, Large sample theory for U-statistics and tests of fit, Ann. Statist. 5 (1977) 110–123.
[10] Y. Guivarc'h, M. Keane and B. Roynette, Marche aléatoires sur les groupes de lie, Lecture Notes in Mathematics 624 (Springer, Berlin, 1977).
[11] P. Hall, Martingale invariance principles, Ann. Probab. 5 (1977) 875–887.
[12] P. Hall, On the Skorohod representation approach to martingale invariance principles, Ann. Probab. 7 (1979) 371–376.
[13] P. Hall, On the invariance principle for U-statistics, Stochastic Process. Appl. 9 (1979) 175–187.
[14] P. Lévy, Wiener's random function and other Laplacian random functions, Proc. 2nd Berkeley Symp. Math. Statist. Probab. (1951)171–187.
[15] P. Lévy, Processus Stochastiques et Mouvement Brownien (Gauthier-Villars, Paris, 1965).
[16] D.L. McLeish, Dependent central limit theorems and invariance principles, Ann. Probab. 2 (1974) 620–628.
[17] D.L. McLeish, Invariance principles for dependent variables, Z. Wahrsch. Verw. Gebiete 32 (1975) 165–178.
[18] D.L. McLeish, On the invariance principle for nonstationary mixingales, Ann. Probab. 5 (1977) 616–621.
[19] R.G. Miller Jr. and P.K. Sen, Weak convergence of U-statistics and von Mises' differentiable statistical functions, Ann. Math. Statist. 43 (1972) 31–41.
[20] G. Morrow and W. Philipp, An almost sure invariance principle for Hilbert space martingales, preprint, 1981.
[21] G. Neuhaus, Functional limit theorems for U-statistics in the degenerate case, J. Multivariate Anal. 7 (1977) 424–439.
[22] H. Rootzén, On the functional central limit theorem for martingales, Z. Wahrsch. Verw. Gebiete 38 (1977) 199–210.
[23] D.J. Scott, Central limit theorems for martingales and for processes with stationary increments using a Skorohod representation approach, Adv. Appl. Probab. 5 (1973) 119–137.
[24] P.K. Sen, Weak convergence of generalized U-statistics, Ann. Probab. 2 (1974) 90–102.